Automatic Modeling of *Personality States* in Small Group Interactions

Jacopo Staiano
DISI-University of Trento
via Sommarive 14
Povo, Italy
staiano@disi.unitn.it

Bruno Lepri
MIT Media Lab and FBK
20 Ames Street,
Cambridge, MA, USA
lepri@fbk.eu

Ramanathan Subramanian
DISI-University of Trento
via Sommarive 14,
Povo, Italy
subramanian@disi.unitn.it

Nicu Sebe
DISI-University of Trento
via Sommarive 14
Povo, Italy
sebe@disi.unitn.it

Fabio Pianesi
FBK
via Sommarive 18
Povo, Italy
pianesi@fbk.eu

ABSTRACT
In this paper, we target the automatic recognition of *personality states* in a meeting scenario employing visual and acoustic features. The social psychology literature has coined the name *personality state* to refer to a specific behavioral episode wherein a person behaves as more or less introvert/extrovert, neurotic or open to experience, *etc*. Personality traits can then be reconstructed as density distributions over personality states. Different machine learning approaches were used to test the effectiveness of the selected features in modeling the dynamics of personality states.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**] [Human information processing]

General Terms: Algorithms, Measurement, Human Factors

Keywords: personality states, modeling, Big Five Traits, modeling dynamics

INTRODUCTION
It is customary for us to describe people as being more or less *talkative*, *bold* or *sociable*. We employ these descriptors in our daily life to explain and/or predict others' behavior, attaching them to well-known and new acquaintances. *Extraversion*, the trait dimension they refer to, is so familiar that we continuously exploit it inconspicuously. Similarly, we talk about other people being more/less prone to anger and frustration (*Neuroticism/Emotional Stability*), responsible or attentive (*Conscientiousness*), and so on. These descriptors relate to traits comprising the well-known Big Five model of personality [1].

The importance of personality for technology and human-computer interaction has also been acknowledged. Studies have shown that personality traits determine people's attitudes towards machines [14] and conversational agents [13]. It has been argued that social networking websites could increase the chances of a successful relationship by analyzing text messages and matching personalities [2], and that tutoring systems would be more effective if they adapt to the learner's personality [7]. Moreover, given its relevance in social settings, information on people's personality can be useful for providing personalized support to group dynamics.

Several works have explored automated personality analysis [9, 11], often targeting the Big Five model of personality [1] of which Extraversion is a major dimension. The general approach is to isolate promising behavioral correlates of the targeted traits for classification or regression, adopting a thin-slice perspective. In particular, some works have focused on the well-known correlation between Extraversion and prosodic features [11] - higher pitch and higher variation of the fundamental frequency, higher voice quality and intensity - while others such as [9], have also considered verbal cues, including many relating to syntax, content, utterance type, *etc*. More recently, Lepri *et al.* [8], exploited medium-grained behaviors enacted in group meetings and related to social attention (social gaze) for automatically predicting *Extraversion*.

All these approaches to the automatic recognition of personality have more or less implicitly adopted the so-called 'person perspective' on personality [3]: for a given behavioral sample, classify whether the sample belongs to an extrovert or introvert (or equivalently, to a neurotic or an emotionally stable, and so on). The problem with this approach is that it assumes a direct and stable relationship between, *e.g.*, being extravert and acting extravertedly (*e.g.*, speaking loudly, being talkative, *etc*.). On the contrary, extraverts can sometimes be silent and reflexive, while introverts can at times exhibit extraverted behaviors. Similarly, people prone to neuroticism need not always exhibit anxious behavior, while agreeable people can sometimes be aggressive. While the person perspective has often dismissed these

* Area Chair: Mor Naaman

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

fluctuations of actual behavior as statistical noise, it has been recently suggested by Fleeson [3] that they are meaningful. The social psychology literature has coined the term *personality states* to refer to concrete behaviors (including ways of acting, feeling and thinking) that can be described as having a similar content to the corresponding personality traits. In other words, a personality state describes a specific behavioral episode wherein a person behaves more or less, introvertly/extravertly, more or less neurotically, *etc*. Personality can then be reconstructed through density distributions over personality states, with parameters such as means, standard deviations, *etc*., summarizing what is specific to the given individual.

In this paper we address the automatic classification of personality states corresponding to the Big Five traits [1], in multi-party meetings. Hence we will be concerned with classifying whether people behaved extravertedly/introvertedly; neurotically or in an emotionally stable manner; in a conscientiousness or careless way; agreeably/disagreeably or creatively/uncreatively. To the best of our knowledge, this is the first work targeting such a qualitative characterization of human behavior in a computational setting, opening new perspectives for the automatic recognition of personality and its relationships to actual behaviors.

THE MISSION SURVIVAL CORPUS

We performed our experiments on meeting videos from the Mission Survival corpus. The Mission Survival II corpus [11] is a multimodal annotated collection of video and audio recordings of meetings in a lab setting. In each meeting, four participants are seated around a table and engaged in the Mission Survival Task (MST), which is used in experimental and social psychology to elicit decision-making in small groups [5]. The objective of the “Mission Survival” task is to reach a consensus on how to survive a disaster scenario, *e.g.*, a plane crash in an uninhabited island. The group has to select a limited number of (up to 15) items critical for survival. Each group member needs to convince the others regarding the utility of a particular item.

The recording equipment consisted of five Fire-wire cameras, four placed in the corners of the room and one directly above the table, and four web-cameras installed on the meeting table. Speech activity was recorded using four close-talk and six table-top microphones along with seven T-shaped microphone arrays, in order to obtain optimal coverage of the environment for speaker localization and tracking.

Personality states

Our experiments were conducted on 4 meeting videos extracted from the Mission Survival corpus (with a total number of 16 subjects and with a total run-time of 120 minutes). The video stream corresponding to each meeting participant was split into 5-minute long slices, making for a total of 108 clips.

Personality state annotation was performed by 30 volunteers (researchers and graduate students) using the Ten Item Personality Inventory [4], a 10-item questionnaire developed to obtain a brief measure of the Big Five dimensions. Annotators were required to assess a participant’s personality based on the 5-minute slices containing a close-up view of the subject with the synchronized audio. Contextual information (the behavior of the other meeting participants) was available only through the audio channel. Each video was annotated by three different annotators, and each annotator saw a given subject no more than once. The wordings of the 10 items were modified in order to reflect the different goal of our exercise; hence, rather than asking to assess how much the item “Extraverted, enthusiastic” applied to the subject, it asked about how much it applied to the behavior he/she exhibited in the given audio-video slice.

In Table 1, we report the global means and standard deviations computed from the annotated personality states. For 16 participants, the average score (5.07 in a range from 1 to 7) for Emotional Stability is quite high, while the mean score for Openness (creativity, complexity) is relatively low (3.95).

In order to perform the classification experiments, scores were quantized (Low/High) for each personality state by taking the median score as a threshold. Table 2 reports the transition probabilities for each personality state. As can be seen, our subjects showed a good degree of consistency in their behaviors: if a person was behaving extravertedly/introvertedly at time t he/she was more likely to continue behave the same way at t+1. The only exceptions in this regard are (a) the transition from a disagreeable to an agreeable behavior; it appears that participants tend to persist in disagreeable behaviors. (b) Also, a transition from an emotionally stable behavior to a neurotic behavior is as likely as continuing with an emotionally stable behavior.

Table 1. Means and standard deviations for personality states

Trait	Mean	Standard deviation
Extraversion	4.18	1.66
Agreeableness	4.55	1.05
Consciousness	4.99	1.12
Em. Stab.	**5.07**	1.04
Openness	**3.96**	1.25

Table 2. Transition probabilities for each personality state

Extraversion	**L**	**H**
L	0.688	0.312
H	0.296	0.704
Agreeableness		
L	0.471	0.529
H	0.259	0.741
Conscientiousness		
L	0.604	0.396
H	0.386	0.614
Emotional Stability		
L	0.500	0.500
H	0.386	0.614
Openness		
L	0.681	0.319
H	0.311	0.689

FEATURE EXTRACTION

A total of 37 features, including both low-level features and high-level features, were automatically extracted from the meeting corpus.

Low-level features

We focused on two classes of features: ‘Activity’ and ‘Emphasis’ [10], measuring vocal signals in social interactions. Activity, implying conversational activity level, usually indicates interest and excitement. It is measured by the z-scored percentage of speaking time (mean and standard deviation of energy per frame, average length in seconds of voiced segments and of speaking segments, fraction of spoken time and voicing rate). For this purpose, the speech stream of each participant is first segmented into voiced and non-voiced segments, and then split into speaking and non-speaking segments. Emphasis is often considered a signal indicating the strength of the speaker's motivation. Moreover, the consistency of emphasis (lower the variations, higher the consistency) could be a signal of mental focus, while variability may signal an openness to influence from other people. Emphasis is measured by the variation in prosody, *i.e.* pitch and amplitude. For each voiced segment, the mean energy, frequency of the fundamental formant and the spectral entropy are extracted (mean of formant frequency, confidence in formant frequency, spectral entropy, value of the largest autocorrelation peak, location of the largest correlation peak, number of the largest autocorrelation peak, time derivative of energy in frame). The mean-scaled standard deviation of these extracted values is then estimated by averaging over longer time-periods (standard deviation of formant frequency, confidence in formant frequency, spectral entropy, value of the largest autocorrelation peaks, location of the largest autocorrelation peaks, number of the largest autocorrelation peaks, and time derivative of energy in frame).

High-level features

Social attention features were extracted by jointly processing the audio-video channels using the hierarchical approach developed in [8]: the output of a Cylindrical Head Model [15] head-pose tracker was used as a proxy for the subject's gaze and fine-tuned with the output of a sub-pixel accurate visual-gaze estimation system [16]. Social attention features were defined as the outcome of joint processing of head-pose and visual-gaze: for each subject, in every frame, *Attention Given* (to at least one of the other participants) and *Attention Received* (from at least one of the other participants) were thus available.

The audio channel was processed through use of a Voice Activity Detection system based on the long-term spectral divergence algorithm detailed in [12]. Speaking time was calculated as the percentage of frames in which voice activity was detected over the duration of the processed slice.

For every participant *p*, the audio-visual features were then fused in order to obtain *Attention Given While Speaking* (the percentage of *p*'s speaking time during which he/she devotes visual attention to at least one of the other members of the group), *Attention Given While Not Speaking* (the percentage of *p*'s non-speaking time during which he/she devotes visual attention to at least one of the other members of the group), *Attention Received While Speaking* (the percentage of *p*'s speaking time during which he/she receives visual attention from at least one of the other group members), and *Attention Received While Not Speaking* (the percentage of *p*'s non-speaking time during which he/she receives visual attention from at least one of the other group members).

FEATURE SELECTION

Correlation-based feature selection is a subset selection technique whose objective is to determine the optimal subset of features [6]. This method evaluates the merit of a subset of features computing the individual predictive ability of each feature along with the degree of redundancy between them. The preferred and selected features using this approach are the features highly correlated with the target value and with low inter-correlation values. This method is used in conjunction with a search strategy, typically the `Best First' search that makes a search in the space of features subsets, using a greedy hill-climbing with a backtracking facility. The search may start with an empty set of features and search forward (forward search) or with the full set of features and search backward (backward search), or at any point search in both directions, forward and backward. We employed the forward search feature selection technique.

The selection has been based on 10-fold cross validation; features selected in at least 8 folds were chosen to train and test the models. The features selected for each personality trait are as follows:

1) *Extroversion* – mean of {attention given, attention received, attention received while not speaking, formant frequency, spectral entropy, energy in frame};
2) *Em. Stab.* – mean of time derivative of energy in frame;
3) *Agreeableness* – mean of {formant frequency, confidence in formant frequency}, standard deviation of {spectral entropy, number of autocorrelation peaks};
4) *Openness* – mean of {attention received, formant frequency, spectral entropy};
5) *Consciousness* – mean of {attention received, attention received while speaking, attention received while not speaking, formant frequency, value of largest autocorrelation peak}.

AUTOMATIC RECOGNITION OF PERSONALITY STATES

A set of machine learning algorithms, both generative (Naïve Bayes, Hidden Markov Models) and discriminative (Support Vector Machines) were used to evaluate the effectiveness of the selected features in modeling the dynamics of personality states. Each algorithm was evaluated in 5 binary classification tasks, one per personality trait. The leave-one-meeting-out strategy was employed, thus 4 models for each personality state were trained on a 3-meetings subset, evaluating them against the remaining one, and finally averaging the results.

The first classifier we applied is Naïve Bayes, a simple probabilistic classifier that applies the Bayes theorem and assumes that the presence/absence of a particular feature of a given class (*e.g.*, a personality state) is unrelated to the presence/absence of other features. The main advantage of using Naïve Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances) necessary for classification. As discussed in connection with Table 2, there are certain regularities in personality state change that a sequential model might be able to capture. Hidden Markov Models were exploited to this end; they consider the temporal correlation between the samples and define the prior probability of the classes in the current sample as depending upon the posterior probabilities of the classes in the previous sample. More precisely, a left-to-right HMM model for each personality state (one for extrovert/introvert; one for neurotic/emotionally stable; and so on) was represented as follows: t, time: $y(t)$, the feature vector; $x(t)$, the personality state; $p(x)$, the priors for the personality states; $p(x(t)|x(t-1)$, the personality states transitions probabilities; $p(y(t)|x(t))$, the conditional distribution of the observed feature vector given the current personality state at time t. We assumed speaker independence and the feature sequences (one per subject) from all the four meetings were used to train a single HMM. The training was done using the standard Expectation Maximization (EM) algorithm. For prediction, each person is represented by an independent instantiation of the same Markov process. Thus, four independent HMMs represent the four people in a meeting.

For classification, we used the standard Viterbi algorithm to compute the most likely sequence of personality states. Finally, we also tested the performance of a powerful discriminative approach. In particular, we used Support Vector Machines (SVMs) as classifiers. The bound-constrained SVM classification algorithms with a linear and with a RBF kernel were used. The cost parameter *C* and the RBF kernel parameter γ were estimated through the grid technique by means of 10-fold cross validation. Furthermore, the cost parameter *C* was weighted for each class with a factor inversely proportional to the class size.

RESULTS AND DISCUSSION

Table 3 reports accuracy results. Bold figures identify values that are significantly better ($p<.05$) than the relevant criteria, represented by the classifiers that use the priors (binomial tests with Bonferroni correction).

Table 3. Summary of classification results

	NB	HMM	SVM (lin)	SVM (rbf)
Extra	**0.694**	**0.731**	**0.648**	**0.676**
En. Stab.	**0.639**	0.574	**0.63**	**0.62**
Agreable	0.583	0.481	0.574	0.583
Opennes	0.5	0.5	0.546	0.547
Conscien	**0.583**	0.547	0.575	0.555

Extraversion is the best-recognized behavioral quality: extravert/introvert behaviors are easier (using the considered features) to distinguish than, *e.g.*, Openness. is the second best recognized quality, an interesting result considering that it is obtained by using just one feature, the (mean of the) speech energy derivative. The results for the other states are either non-significant or barely significant (as for Conscientiousness with Naïve Bayes).

HMMs yield interesting results only with Extraversion and in this case, they perform statistically better than linear or rbf SVMs. Since in the ground truth, personality states such as Openness and Conscientiousness show temporal properties similar to those of Extraversion, the ineffectiveness of HMM to isolate them is probably due to the limited predictive power of the selected features. Finally, SVM classifiers with linear and RBF kernels have similar performance scores.

In general, it seems that the considered non-verbal features have a good power in discriminating between introvert and extrovert behaviors, a result that emphasizes the important role of vocal (*e.g.*, pitch) and social attention features not only for recognizing the Extroversion trait (as widely documented by studies in social psychology and in the automatic analysis of behavior), but also for the recognition of behaviors expected from the trait. The encouraging results obtained for the Neurotic dimension, employing a single feature are also notable. Putting them together with the results concerning Extraversion, and noting that the two corresponding traits are also referred to as Positive and Negative Affect respectively, it can be concluded that the quality of affect, as it surfaces in actual behaviors, is a most readily detectable characteristic among those constituting the Big Five taxonomy.

We have just begun work concerning a highly complex phenomenon: automatic characterization of the quality of behaviors in terms of personality states. Still, preliminary results appear to have proven its feasibility, opening the way to a new and exciting research area.

AKNOWLEDGEMENT

Bruno Lepri's research is funded by PERSI project inside the Marie Curie COFUND – 7th Framework. The work of Jacopo Staiano, Ramanathan Subramanian, and Nicu Sebe was partially supported by the FIRB S-PATTERNS project.

REFERENCES

Costa, P.T., and McCrae, R.R. Four ways why five factors are basic. *Personality and Individual Differences*, 13(6): 653-665, 1992.

Donnellan, M.B., Conger, R.D., and Bryant, C.M. The Big Five and enduring marriages. *Journal of Research in Personality*, 38: 481–504, 2004.

Fleeson, W. Towards a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80: 1011-1027, 2001.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality*, 37: 504-528, 2003.

Hall J., and Watson W. The effects of a normative intervention on group decision-making performance. *Human Relations*, 23(4): 299–317, 1970.

Hall, M. A. Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning. *Proc. 17th Int'l Conf. Machine Learning*, 359-366, 2000.

Komarraju, M., and Karau, S. J. The relationship between the Big Five personality traits and academic motivation. *Personality and Individual Differences*, 39: 557–567, 2005.

Lepri, B. Subramanian, R., Kalimeri, K., Staiano, J., Pianesi, F., and Sebe, N. Employing Social Gaze and Speaking Activity for Automatic Determination of the Extraversion Trait. In *Int'l Conf. on Multimodal Interfaces*, 2010.

Mairesse F., Walker M.A., Mehl M.R., and Moore R.K. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial Intelligence Research* 30: 457-500, 2007.

Pentland, A. *Honest Signals*. MIT Press, 2008.

Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., and Zancanaro. M. Multimodal Recognition of Personality Traits in Social Interactions. In *Int'l Conf. on Multimodal Interfaces*, 2008.

Ramirez, J., Segura, C., Benitez, C., de la Torre, A., and Rubio,A.Efficient voice activity detection algorithms using long-term speech information. *Speech Communication*. 42(3-4): 271-287, 2004.

Reeves, B., and Nass, C. *The Media Equation*. University of Chicago Press, 1996.

Sigurdsson, J.F. Computer experience, attitudes toward computers and personality characteristics in psychology undergraduates. *Personality and Individual Differences*, 12(6): 617–624, 1991.

Valenti, R., Sebe, N., and Gevers, T. Combining Head Pose and Eye Location Information for Gaze Estimation. In *IEEE Trans. Image Processing*, 2011.

Valenti, R., Staiano, J., Sebe, N., and Gevers, T. Webcam-based Visual Gaze Estimation. In *Int'l Conference on Image Analysis and Processing*, 662–671, 2009.